Experimental demonstration of counterfactual quantum key distribution

Min Ren, Guang Wu,* E Wu, and Heping Zeng†
State Key Laboratory of Precision Spectroscopy, East China Normal University, Shanghai 200062, China

(Dated: March 25, 2010)

Counterfactual quantum key distribution provides natural advantage against the eavesdropping on the actual signal particles. It can prevent the photon-number-splitting attack when a weak coherent light source is used for the practical implementation. We realized the counterfactual quantum key distribution in an unbalanced Mach-Zehnder interferometer of 12.5-km-long quantum channel with a high-fringe visibility of 96.4%. As a result, we obtained secure keys against the noise-induced attack (eg. the vacuum attack) and passive photon-number-splitting attack.

PACS numbers: 03.67.Dd

Quantum key distribution (QKD) provides an unconditionally secure communication between two remote parties (Alice and Bob), where the security is guaranteed by the fundamentals of quantum mechanics [1, 2]. However, current techniques cannot support to implement the ideal quantum cryptography experiment as originally proposed due to the lack of efficient single-photon or entangled photon-pair sources in the near-infrared region. So far, in practical long-distance fiber-based QKD systems, weak coherent light sources have been used instead of ideal single photons [3–5]. Intelligent methods have been invented to prevent the photon-number-splitting (PNS) attack based on the multi-photon pulse of the weak coherent light source [6–9]. Among them, the decoy-state QKD protocol [10–12] could prevent the PNS attack by statistical security analysis of a large yield of the photon clicks [13, 14]. On the other hand, quantum counterfactual effect has been discussed in interaction-free measurement [15–17] and developed for the quantum computation [18–20]. Recently, an interesting counterfactual QKD protocol was proposed [21] to distribute secret keys without any secret information-carrier qubits transmitting in the quantum channel, which can in principle prevent the eavesdropper to directly access the entire quantum system of each qubit. For example, secret key can be established by using the quantum counterfactual effect in a Michelson interferometer between Alice and Bob. One arm of the interferometer is in Alice's secure site, and the other arm is used to connect Bob as the quantum channel. Unlike previous QKD protocols which generate secret keys by transmitting the signal particles through the actual quantum channel, the proposed counterfactual QKD generates secret keys in the case that the photonic qubits only transmit in Alice's arm but never travel through the quantum channel to Bob. Interestingly, as a weak coherent light source is used, the quantum counterfactual effect provides a natural advantage to prevent the PNS attack by avoiding Eve's access to the information-carrier photons. Experimental demonstration of such a counterfactual QKD requires a stabilized Michelson interferometer [22].

In this letter, we experimentally demonstrate the counterfactual QKD based on a round-way unbalanced Mach-Zehnder interferometer with 25 km fiber length difference between the long and short arms that ensured an effective 12.5-km-long quantum channel between Alice and Bob. Possible eavesdropping against the counterfactual QKD was analyzed. The counterfactual QKD scheme was implemented with the capability to reveal the vacuum attack by monitoring the photon detection distributions. With a high-fringe visibility of 96.4%, we could obtain secure keys against the passive PNS attack by an average photon number of 1 per pulse.

The counterfactual QKD system was composed of a round-way unbalanced Mach-Zehnder interferometer as shown in Fig. 1. A 1550-nm DFB laser diode (LD) generated a series of short pulses with the repetition rate of 5 kHz, which were adjusted to horizontal polarization at the input of the interferometer. The laser pulses were re-

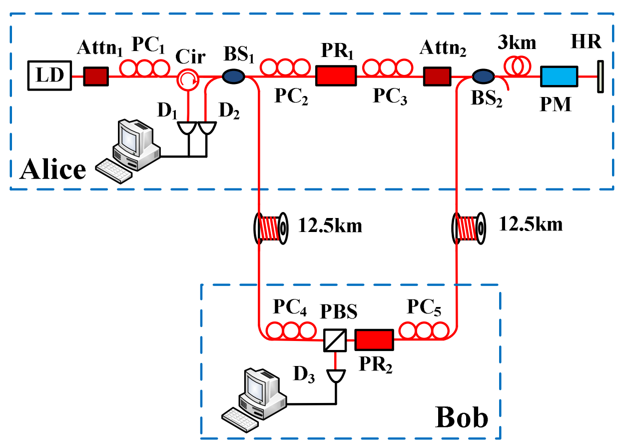

FIG. 1: Schematic of the counterfactual QKD system. LD: 1550-nm DFB laser diode; $\mathrm{Attn}_{1,2}$: variable attenuators; PC_{1-5}: manual polarization controllers; Cir: circulator; $\mathrm{BS}_{1,2}$: 50:50 beam splitters; $\mathrm{PR}_{1,2}$: polarization rotators; PM: phase modulator; HR: high-reflection mirror; PBS: polarization beam splitter; D_{1-3}: single-photon detectors.

*Electronic address: gwu@phy.ecnu.edu.cn
†Electronic address: hpzeng@phy.ecnu.edu.cn

flected by a high-reflection mirror (HR), and attenuated to single-photon level before returning the interferometer. A pulse from the laser source might travel one of the four paths by the selection of the short arm (s) or long arm (l) in the forward and backward transmissions, denoted as (ss), (sl), (ls), and (ll), respectively. We disregarded the photon paths ll and ss in the counterfactual QKD, since ll states contained no information, and ss states were in Alice's secure station with no information leak. Only the photons travelling sl and ls paths contributed to QKD, and the interference was automatically stabilized since photons travelled the fiber paths of exactly the same length. Four manual polarization controllers (PC_{2-5}) were adjusted for compensation of the polarization drifts in the long-distance fiber to make the photons of sl and ls paths interfere with the same polarizations at the output of the static interferometer.

The static interferometer was then used to polarization-encode the qubit in both Alice's and Bob's sites with two polarization rotators (PR_1 and PR_2), which were randomly controlled to rotate the polarization state by either 0° or 90°. In the forward transmission, Alice randomly modulated the laser pulses passing the short arm to the horizontal or vertical polarization direction. Then two possible quantum states of the photon pulses were prepared as

$$\begin{aligned} |\psi_0\rangle &= (|0\rangle_{sl}|H\rangle_{ls} + |H\rangle_{sl}|0\rangle_{ls})/\sqrt{2}, \\ |\psi_1\rangle &= (|0\rangle_{sl}|H\rangle_{ls} + |V\rangle_{sl}|0\rangle_{ls})/\sqrt{2}, \end{aligned} \tag{1}$$

where $|0\rangle$ is the vacuum state, $|H\rangle$ and $|V\rangle$ denote the horizontal and vertical polarization states, defined as the bit value of 0 and 1, respectively. In the backward transmission, Bob decoded the polarization states with the polarization rotator (PR_2) and the polarization beam splitter (PBS) by randomly applying a 90° or 0° polarization rotation on the split pulse in the long arm, corresponding to the bit value of 0 or 1. If Alice and Bob chose the same bit value, the split pulse in the long arm was reflected to the single-photon detector D_3 by the polarization beam splitter. The quantum states $|\psi_0\rangle$ and $|\psi_1\rangle$ both collapsed to $|V\rangle_{sl}|0\rangle_{ls}$ or $|0\rangle_{sl}|H\rangle_{ls}$. The photon travelled either the short arm, which was then detected by the single-photon detector D_1 or D_2 with the same probability of 1/4, or the long arm, which was then detected by the single-photon detector D_3 with the probability of 1/2. If Alice and Bob chose different bit values, the split pulse in the long arm passed through the polarization beam splitter. The quantum state kept the interference, and the photon pulses went toward the single-photon detector D_1, wherein $|\psi_0\rangle$ was unchanged, and $|\psi_1\rangle$ was transformed to $|\psi_0\rangle$ since the polarization of the split pulse in the long arm was rotated by 90°.

Only the events that D_2 alone detected a photon would create the sifted keys, indicating that Alice and Bob have certainly chosen the same bit value and the photon pulses travelled the short arm. If D_1 or D_3 clicked, Alice and Bob announced the detected results and their bit values. The performance of the interference could be calculated from these events to monitor whether there were any Eve's disturbance. We defined ($C1$, $C3$, $C5$) and ($C2$, $C4$, $C6$) as the counts of the detectors (D_1, D_3, D_2) in the case that Alice and Bob chose the same and different bit values, respectively. The count of D_2 was randomly selected from the half of the events that D_2 clicked, and the other half events created the sifted keys. The performance of the interference and polarization extinction ratio of Alice's and Bob's polarization operation could be respectively characterized by $C1 : C2$ and $C3 : C4$, where $C1 : C2 = 1 : 4$ represents an ideal interference [23]). The interference fringe visibility could be calculated by $(4C5 - C6)/(4C5 + C6)$, while the error rate of the QKD system could be calculated by $C5/(C5 + C6)$.

As Eve had no access to the short arm and the information-carrier photons, many possible attacks that threaten the conventional QKD, such as intercept-resent and Trojan horse attacks, could be successfully prevented [24]. Eve may implement a possible eavesdropping by using "vacuum attack", in which Eve operates the polarization-encoding on the vacuum state before Bob's site. Eve's choice of bit value (B_{Eve}) may be the same or different with Alice's (B_{Alice}) or Bob's (B_{Bob}). There are altogether four possible cases. (i) $B_{Eve} = B_{Alice} = B_{Bob}$, Eve may eavesdrop the sifted key with no errors for Alice and Bob. (ii) $B_{Eve} = B_{Alice} \neq B_{Bob}$, the quantum state collapses, while Eve may eavesdrop the sifted key and induce an error for Bob. (iii) $B_{Eve} \neq B_{Alice} = B_{Bob}$, the photon keeps the interference and is detected by D_1. (iv) $B_{Eve} \neq B_{Alice} \neq B_{Bob}$, the quantum state collapses, while Eve may eavesdrop a sifted key with a possible error. Three of the four cases may create a sifted key, wherein Bob and Eve both have the error rate of 1/3 in the sifted key. According to the Shannon information theory, the information Eve and Bob get is described as

$$I = 1 + Dlog_2(D) + (1 - D)log_2(1 - D), \tag{2}$$

where D is the error rate. Here Eve's information through the vacuum attack is only 8.17%, which is much less than 50% in the conventional QKD through the intercept-resent attack. Moreover, the vacuum attack changes the count of D_1 and D_2, as well as $C1 : C2 \longrightarrow 1 : 2$ [23].

The counterfactual cryptography naturally prevents the PNS attack when a weak coherent light source is used, because Eve even cannot detect the photon number of each pulse in the long arm, and she has no access to the short arm [21]. Passive PNS attack would provide Eve some information by inserting a beam splitter in the quantum channel to passively split the photon and replacing the quantum channel with her lossless channel to compensate the splitting loss. Eve's gain of the sifted key (P_{Eve}) is extremely limited because only the joint event that there is at least one photon in each interferometer arm is useful in this attack. The proportion of Eve's information to the total sifted key is given by

$$P_{Eve} \leq P(\mu_{sl}, n > 0) \cdot \eta_{loss}, \tag{3}$$

where μ_{sl} is the mean photon number of the photon pulse in the long arm, μ is the mean photon number of $|\psi_0\rangle$ or $|\psi_1\rangle$. Here $\mu_{sl} = 0.5 \cdot \mu$, if the long and short arms have the same attenuation. $P(\mu, n) = e^{-\mu}\mu^n/n!$ is the Poissonian distribution for the photon number of n. $\eta_{loss} = 1 - 10^{-0.25L_{eff}/10}$ is the transmission loss of the quantum channel from BS_2 to PC_3, where L_{eff} refers to the effective distance between Alice and Bob. When the loss of the quantum channel is so large that $\eta_{loss} \to 1$, Eve could nearly obtain all the information of the photon in the long arm, and her information gain won't increase with the communication distance any more.

According to the Shannon information theory, the secure key can be distilled if Bob obtains more information than Eve. As shown in Fig. 2, Eve's maximum gain of the sifted key was less than 5% for the case $\mu = 0.1$, independent on the large loss of the long-distance channel. As a result, the counterfactual QKD protocol provides an approach to the long-distance secure communication based on the quantum property instead of the statistical method such as decoy-state protocols.

In the experiment, the total loss of the long arm was about 9 dB, while the loss of the short arm was only 3 dB. The photon flux in the long arm was larger than that in the short arm in the backward transmission, resulting in an increase of Eve's gain. To eliminate the threaten induced by the unbalance of the two arms, we inserted an attenuator ($Attn_2$), and adjusted the total loss of the short arm to be the same as that of the long arm. Three home-made near-infrared single-photon detectors were used with the dark-count probabilities of 1.0×10^{-5}, 5×10^{-6}, 1.5×10^{-5} per pulse, respectively, at a detection efficiency of 10%. Due to the birefringence of the single-mode fiber, the photon pulses of ls or sl path would cover a different phase shift when Alice sent the horizontal or vertical polarization state. As there was about 0.4π phase shift between two orthogonal polarization in 12.5 km single-mode fiber, we used a phase modulator to compensate the polarization-dependent phase shift. The laser pulses were attenuated to contain 1 photon per pulse on average. Then we measured the single-photon interference and the polarization extinction ratio of the quantum system. As shown in Fig. 3, when Alice and Bob chose the same or different polarization, the photon pulses in the long arm were switched to D_3, or passed through the polarization beam splitter to the output of the interferometer. The photon count of D_3 showed a polarization extinction ratio as high as $30:1$. The single photon interfered at the output of the interferometer when Alice and Bob chose different polarization. Owing to the passive phase shift compensation of the round-way Mach-Zehnder interferometer, the phase shift caused by the slow variation such as random temperature and stress drifts were auto-compensated. In this way, we got a stable interference with the fringe visibility of 96.4%. When Alice and Bob chose the same polarization, the quantum state of the signal photons collapsed, and we got almost the same count rates on D_1 and D_2.

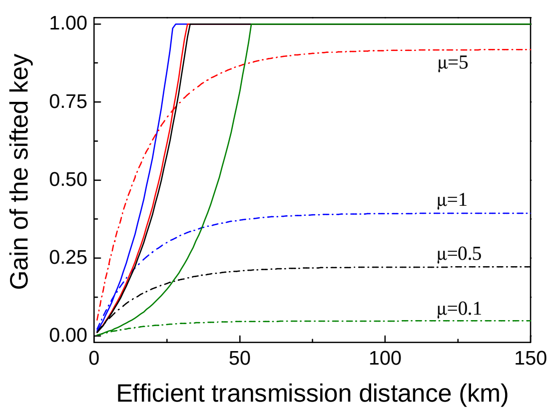

FIG. 2: Eve's gain of the sifted key. Dashed and solid lines show Eve's gain through passive PNS attack against the counterfactual and PNS attack against conventional QKD, respectively. The olive, black, blue, and red lines denote the mean photon number of 0.1, 0.5, 1.0, and 5.0, respectively.

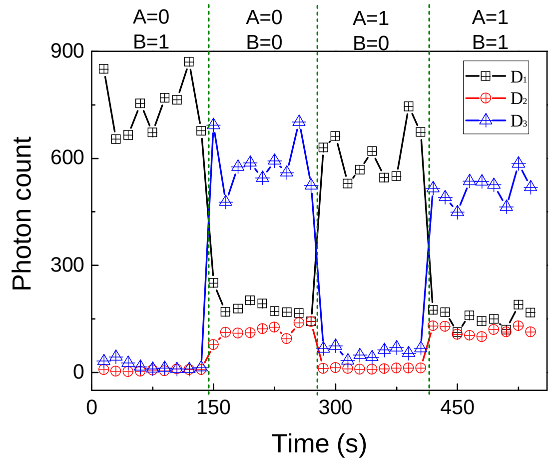

FIG. 3: Photon counts of D_1, D_2 and D_3 at various bit value choice of Alice and Bob with the acquisition time of 15 s.

The experimental demonstration of the counterfactual QKD was then carried out with the average photon number $\mu = 1.0$. Alice and Bob used two random number generators to drive PR_1 and PR_2 independently. Table I presents the experimental results with an acquisition time of 540 s. The whole system showed an unsurpassed stability during the acquisition. The counts of each single-photon detector and the coincidence counts between D_3 and D_1 or D_2 showed that the photon distribution experienced no observable changes. The error rate D_{AB} was 6.7%, mainly from the noise of the interference. Note that the counterfactual QKD requires a single-photon interference of a high-fringe visibility, which would have fourfold errors as the conventional QKD of the same fringe visibility. The interference errors were mainly induced by the polarization operation.

TABLE I: Experimental results of 12.5-km counterfactual QKD with the acquisition time of 540 s.

D_1	15149
D_2	2243
D_3	10577
$C1 : C2$	1 : 4.1
Fringe Visibility	96.4%
Polarization Extinction Ratio	30.5 : 1
Coincidence Probability	0.15%
Sifted key	1121
D_{AB}	6.7%

The polarization extinction ratio was kept at 30.5 : 1 in the experiment. This error rate opened the back door for Eve to implement the vacuum attack against 20% of the vacuum states. However, it would change the counts of D_1 and D_2, and $C1 : C2 \longrightarrow 1 : 3.3$ [23]. In the experiment, we got $C1 : C2 = 4.1$, implying that the error rate didn't come from the vacuum attack but was induced by the intrinsic noise of the QKD system. In this case, Eve may obtain 20% of the information through the passive PNS attack, while Bob's information was 64.5% according to Eq. 2. Note that Eve's information was restricted under a certain level under this attack (Fig. 2) and was insensitive to the transmission loss. Since Bob's information was much more than Eve's, Alice and Bob could distill the unconditional secure keys at last by the classical error correction and privacy amplification methods [2, 25, 26].

In conclusion, we realized the counterfactual QKD experiment in a round-way unbalanced Mach-Zehnder interferometer of 25 km fiber length difference between the long and short arms that ensured a fiber-based quantum channel of 12.5 km. Despite Bob had an error rate of 6.7%, we could ensure that there was no noise-induced attack according to the unchanged count distribution of each single-photon detectors used in the QKD system, and secure keys could be obtained against the passive PNS attack. The key generation rate of the counterfactual QKD, which was mainly limited by the slow response time of the polarization rotators used in our experiment, could be increased with fast polarization rotators and high-speed single-photon detectors [27, 28]. And the active polarization compensating methods have been invented for long-distance fiber-based QKD experiments[29–31], which were quite useful to realize a long-term stable and long-distance counterfactual QKD system. The counterfactual QKD scheme was implemented with currently available technologies, promising a robust and practical quantum cryptography system toward global secure communication.

This work was funded in part by National Natural Science Fund of China (10525416, 10904039 and 10990101), and National Key Project for Basic Research (2006CB921105).

[1] C. H. Bennett and G. Brassard, in Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India (IEEE, New York, 1984), p. 175.

[2] N. Gisin,G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).

[3] C. Gobby, Z. L. Yuan, and A. J. Shields, Appl. Phys. Lett. **84**, 3762 (2004).

[4] X. F. Mo, B. Zhu, Z. P. Han, Y. Z. Gui, and G. C. Guo, Opt. Lett. **30**, 2632 (2005).

[5] C. Z. Peng, J. Zhang, D. Yang, et. al., Phys. Rev. Lett. **98**, 010505 (2007).

[6] A. Acin, N. Gisin, and V. Scarani, Phys. Rev. A **69**, 012309 (2004).

[7] M. Koashi, Phys. Rev. Lett. **93**, 120501 (2004).

[8] O. L. Guerreau, F. J. Malassenet, S. W. McLaughlin, and J. M. Merolla, IEEE Photon. Tech. Lett. **1**, 1 (2005).

[9] G. Wu, J. Chen, Y. Li, L. L. Xu, and H. P. Zeng, Phys. Rev. A **74**, 062323 (2006).

[10] W. Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003).

[11] X. B. Wang, Phys. Rev. Lett. **94**, 230503 (2005)

[12] H. K. Lo, X. F. Ma, and K. Chen, Phys. Rev. Lett. **94**, 230504 (2005).

[13] X. F. Ma, B. Qi, Y. Zhao, and H. K. Lo, Phys. Rev. A **72**, 012326 (2005).

[14] X. B. Wang, Phys. Rev. A **72**, 012322 (2005).

[15] A. C. Elitzur and L. Vaidman, Found. Phys. **23**, 987 (1993).

[16] P. Kwiat, H. Weinfurter, T. Herzog, A. Zeilinger, and M. A. Kasevich, Phys. Rev. Lett. **74**, 4763 (1995).

[17] R. Penrose, Shadows of the Mind (Oxford Univ. Press, New York, 1994), p. 240.

[18] R. Jozsa, in Lecture Notes in Computer Science, edited by C. P. Williams (Springer-Verlag, Berlin, 1999), Vol. 1509, p. 103.

[19] G. Mitchison and R. Jozsa, Proc. R. Soc. A **457**, 1175 (2001).

[20] O. Hosten, M. T. Rakher, J. T. Barreiro, N. A. Peters, and P. G. Kwiat, Nature (London) **439**, 949 (2006).

[21] T. G. Noh, Phys. Rev. Lett. **103**, 230501 (2009).

[22] S. B. Cho and T. G. Noh, Opt. Express **17**, 19027 (2009).

[23] If there is no eavesdropping, the qubit collapses or interferes with the same probability (D_1 clicks certainly). So $C1 = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$, and $C2 = \frac{1}{2} \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{2}$. If Eve does the vacuum attack, the qubit collapses (or interferes) with the probability of $\frac{3}{4}$ (or $\frac{1}{4}$). So $C1 = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{16}$, and $C2 = \frac{1}{2} \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{4} \cdot 1 = \frac{3}{8}$.

[24] T. G. Noh, arXiv:0809.3979v2.

[25] C. H. Bennett, G. Brassard, and J.-M. Robert, SIAM J. Comput. f 17, 210C229 (1988).

[26] G. Brassard and L. Salvail, in Advances in Cryptology, EUROCRYPT'93 Proceedings, Lecture Notes in Computer Science, Vol. 765, edited by T. Helleseth (Springer, New York, 1994), p. 410.

[27] Z. L. Yuan, B. E. Kardynal, A. W. Sharpe, and A. J.

Shields, Appl. Phys. Lett. 91, 041114 (2007)
[28] L. L. Xu, E Wu, X. R. Gu, Y. Jian, G. Wu, and H. P. Zeng, Appl. Phys. Lett. 94, 161106 (2009).
[29] J. Chen, G. Wu, Y. Li, E Wu, and H. P. Zeng, Opt. Express **15**, 17928 (2007).
[30] G. B. Xavier, G. Vilela de Faria, G. P. Temporão, and J. P. von der Weid, Opt. Express **16**, 1867 (2008)
[31] J. Chen, G. Wu, L. L. Xu, X. R. Gu, E Wu, and H. P. Zeng, New J. Phys. **11**, 065004 (2009).